Wykład 1: Zmienne losowe i rozkłady prawdopodobieństwa. Dyskretne i ciągłe rozkłady służące do modelowania

Do opisu zjawisk losowych służą *zmienne losowe*, czyli funkcje mierzalne, które zdarzeniom elementarnym przyporządkowują liczby rzeczywiste. Zmienne losowe opisujemy poprzez podanie ich rozkładu prawdopodobieństwa. Podstawowe typy zmiennych losowych (i zarazem rozkładów prawdopodobieństwa) to:

1. zmienne losowe dyskretne (i odpowiadające im dyskretne rozkłady prawdopodobieństwa);
2. zmienne losowe absolutnie ciągłe (i odpowiadające im absolutnie ciągłe rozkłady prawdopodobieństwa).

Zmienne losowe dyskretne (inaczej zwane **skokowymi)** przyjmują skończoną bądź przeliczalną liczbę wartości. Rozkład tych o skończonej liczbie wartości wygodnie jest opisywać za pomocą tabelki.

ROZKŁADY DYSKRETNE

1). Rozkład dwupunktowy: Wykonujemy doświadczenie, które może zakończyć się jedynie sukcesem (z prawdopodobieństwem $p \in (0,1)$) albo porażką. Zmienna losowa X przyjmuje wartość 1, gdy zajdzie sukces, oraz wartość 0, gdy zajdzie porażka. Wtedy rozkład X opisuje poniższa tabelka:

x_k	0	1
p_k	$1-p$	p

,

co równoważnie można zapisać następująco:

$$P(X=0) = 1-p \ \text{ i } \ P(X=1) = p.$$

2). Rozkład dwumianowy: $X \sim binom(n,p)$.

Powtarzamy n-krotnie doświadczenie, które może zakończyć się jedynie sukcesem (z prawdopodobieństwem $p \in (0,1)$) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba uzyskanych sukcesów.

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{dla } k = 0, 1, \ldots, n.$$

3). Rozkład Poissona: $X \sim Pois(\lambda)$, gdzie $\lambda > 0$, jeśli

$$P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{dla } k = 0, 1, \ldots.$$

Twierdzenie (Poissona)**.** Jeśli $X_1, X_2, X_3, \ldots$ jest ciągiem zmiennych losowych takim, że $X_n \sim binom(n, p_n)$, gdzie $np_n \overset{n\to\infty}{\longrightarrow} \lambda$, to

$$\lim_{n\to\infty} P(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{dla } k = 0, 1, \ldots.$$

Rozkład Poissona może posłużyć do modelowania, np. liczby cząstek emitowanych przez daną substancję promieniotwórczą w ciągu 10 sek.; liczby klientów odwiedzających dany oddział banku w ciągu godziny; liczby awarii pewnego urządzenia w ciągu dnia; liczby wypadków na danym skrzyżowaniu w ciągu tygodnia.

4). Rozkład geometryczny: $X \sim Ge(p)$, gdzie $p \in (0,1)$.

Powtarzamy doświadczenie, które może zakończyć się sukcesem (z prawdopodobieństwem p) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba porażek poprzedzających pierwszy sukces.

$$P(X = k) = p(1-p)^k \ \text{ dla } \ k = 0, 1, \ldots.$$

5). Rozkład ujemny dwumianowy: $X \sim nb(p, r)$, gdzie $p \in (0,1)$, $r \in \{1, 2, \ldots\}$.

Powtarzamy doświadczenie, które może zakończyć się sukcesem (z prawdopodobieństwem p) albo porażką. Zakładamy, że kolejne powtórzenia doświadczenia przebiegają niezależnie od pozostałych. Wtedy X to liczba porażek poprzedzających r-ty sukces.

$$P(X = k) = \binom{r+k-1}{k} p^r (1-p)^k \ \text{ dla } \ k = 0, 1, \ldots.$$

Rozkład geometryczny i ujemny rozkład dwumianowy mogą posłużyć do modelowania, np. dyskretnych czasów działania urządzeń (w szczególnoci do modelowania liczby włączeń i wyłączeń danego urządzenia do momentu jego popsucia się czy liczby pełnych dni przepracowanych przez dane urządzenie do czasu awarii).

Zmienne losowe (absolutnie) ciągłe przyjmują nieprzeliczalną liczbę wartości i ich rozkład można opisać przez podanie *funkcji gęstości* $f(x)$ czyli nieujemnej funkcji $f : \mathbb{R} \to \mathbb{R}$ takiej, że

$$P(X \in A) = \int_A f(x)dx \qquad \text{dla dowolnego zbioru mierzalnego } A.$$

Zauważmy, że dla zmiennych losowych ciągłych mamy

$$P(X = a) = \int_{\{a\}} f(x)dx = 0,$$

pomimo, że zdarzenie $\{X = a\}$ nie musi być niemożliwe.

PRZYKŁADY ROZKŁADÓW CIĄGŁYCH:

1). Rozkład jednostajny na przedziale (a, b), oznaczany $U(a, b)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{dla} \quad x \in (a, b) \\ 0 & \text{dla} \quad x \notin (a, b). \end{cases}$$

2). Rozkład wykładniczy z parametrem $\lambda > 0$, oznaczany $Exp(\lambda)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla} \quad x > 0 \\ 0 & \text{dla} \quad x \leqslant 0. \end{cases}$$

Rozkład wykładniczy może być adekwatny do modelowania czasów oczekiwania na pierwszy sukces, np. czasu oczekiwania na przyjście pierwszego klienta do sklepu lub czasu oczekiwania na emisję pierwszej cząstki przez daną substancję promieniotwórczą.

Może on także posłużyć do modelowania ciągłych czasów życia obiektów. Jednak trzeba pamiętać, że ma on tzw. własność braku pamięci, tzn. jeśli $X \sim Exp(\lambda)$, to

$$P(X > x + y | X > x) = P(X > y) \ \ \textit{dla } x > 0, y > 0.$$

Jeśli modelowany czas nie ma tej własności, to powinniśmy użyć innego rozkładu. Wtedy przydatny może okazać się rozkład Weibulla.

3). <u>Rozkład Weibulla</u> z parametrem kształtu a i parametrem skali b, oznaczany $Weibull(a,b)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}, \ a > 0, b > 0.$$

Rozkład $Weibull(a,b)$ z $a = 1$ to po prostu rozkład wykładniczy $Exp(1/b)$, zatem ma on własność braku pamięci.

Natomiast dla rozkładu $Weibull(a,b)$ z $a > 1$ awaryjność rośnie wraz z upływem czasu, zaś dla rozkładu $Weibull(a,b)$ z $a < 1$ awaryjność maleje wraz z upływem czasu, gdzie przez awaryjność rozumiemy funcję

$$R(t) = P(X < t + y | X > t), \ t > 0, y > 0.$$

4). <u>Rozkład gamma</u> z parametrem kształtu a i parametrem skali s, oznaczany $Gamma(a,s)$, czyli rozkład o gęstości

$$f(x) = \begin{cases} \frac{1}{s^a \Gamma(a)} x^{a-1} \exp\left(-x/s\right) & \text{dla } x > 0 \\ 0 & \text{dla } x \leqslant 0 \end{cases}, \ a > 0, s > 0,$$

gdzie $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ (dla $n = 1, 2 \ldots$ mamy $\Gamma(n) = (n-1)!$).

Rozkład gamma może posłużyć do modelowania wielkości opadów w danym miejscu lub do modelowania wysokości roszczenia ubezpieczeniowego.

5). <u>Rozkład normalny</u> o parametrach μ i σ^2, oznaczany $\mathcal{N}(\mu, \sigma^2)$ – zostanie omówiony w następnym wykładzie.

WIELKOŚCI SŁUŻĄCE DO OPISU ZMIENNYCH LOSOWYCH:

1). Dystrybuanta : $F(x) = P(X \leqslant x)$.

2). Funkcja przeżycia : $\overline{F}(x) = 1 - F(x) = P(X > x)$.

3). Funkcja kwantylowa (jest to funkcja odwrotna do dystrybuabty, jeśli tylko takowa istnieje):

$$q(\alpha) \text{ dla } \alpha \in (0,1) \text{ to najmniejsza liczba spełniająca } F(q(\alpha)) \geqslant \alpha.$$

4). Wartość oczekiwana :

$$EX = \begin{cases} \sum_k x_k p_k & \text{dla zmiennych losowych dyskretnych} \\ \int_{\mathbb{R}} x f(x) dx & \text{dla zmiennych losowych ciągłych.} \end{cases}$$

Własności wartości oczekiwanej: jeśli $a \in \mathbb{R}$, to

- $E(a) = a$,
- $E(aX) = aE(X)$,
- $E(X \pm Y) = EX \pm EY$.

5). Wariancja : $Var(X) = E(X^2) - (EX)^2$, gdzie

$$EX^2 = \begin{cases} \sum_k x_k^2 p_k & \text{dla zmiennych losowych dyskretnych} \\ \int_{\mathbb{R}} x^2 f(x) dx & \text{dla zmiennych losowych ciągłych.} \end{cases}$$

Własności wariancji: jeśli $a \in \mathbb{R}$, to

- $Var(a) = 0$,
- $Var(aX) = a^2 Var(X)$;
- jeśli X i Y są niezależne, to $Var(X \pm Y) = Var(X) + Var(Y)$;
- $Var(X) \geqslant 0$.

6). Odchylenie standardowe : $\sigma_X = \sqrt{Var(X)}$.

Dla przykładowych rozkładów zmiennych losowych mamy:

- jeśli $X \sim binom(1,p)$, to $EX = p$ i $Var(X) = p(1-p)$;
- jeśli $X \sim binom(n,p)$, to $EX = np$ i $Var(X) = np(1-p)$;
- jeśli $X \sim Pois(\lambda)$, to $EX = Var(X) = \lambda$;
- jeśli $X \sim Ge(p)$, to $EX = \frac{1-p}{p}$ i $Var(X) = \frac{1-p}{p^2}$;
- jeśli $X \sim nb(p,r)$, to $EX = \frac{(1-p)r}{p}$ i $Var(X) = \frac{(1-p)r}{p^2}$;
- jeśli $X \sim U(a,b)$, to $EX = \frac{a+b}{2}$ i $Var(X) = \frac{(a+b)^2}{12}$;
- jeśli $X \sim Exp(\lambda)$, to $EX = \frac{1}{\lambda}$ i $Var(X) = \frac{1}{\lambda^2}$;
- jeśli $X \sim Weibull(a,b)$, to $EX = b\Gamma(1+\frac{1}{a})$ i $Var(X) = b^2\left(\Gamma(1+\frac{2}{a}) - \left(\Gamma(1+\frac{1}{a})\right)^2\right)$;
- jeśli $X \sim Gamma(a,s)$, to $EX = as$ i $Var(X) = as^2$;
- jeśli $X \sim \mathcal{N}(\mu,\sigma^2)$, to $EX = \mu$ i $Var(X) = \sigma^2$.

FUNKCJE W PAKIECIE R:

rozkład	$P(X=x)$ dla rozkładów dyskretnych lub $f(x)$ dla rozkładów ciągłych
$binom(n,p)$	dbinom(x, size=n, prob=p)
$Pois(\lambda)$	dpois(x, lambda= λ)
$Ge(p)$	dgeom(x, prob=p)
$nb(p,r)$	dnbinom(x, size=r, prob=p)
$U(a,b)$	dunif(x, min=a, max=b)
$Exp(\lambda)$	dexp(x, rate= λ)
$Weibull(a,b)$	dweibull(x, shape=a, scale=b)
$Gamma(a,s)$	dgamma(x, shape=a, scale=s)
$\mathcal{N}(\mu,\sigma^2)$	dnorm(x, mean= μ, sd= σ)

Ponadto dla rozkładu dwumianowego:

- dystrybuanta: $P(X \leqslant x)$ =pbinom(q=x, size=n, prob=p, lower.tail=TRUE);
- ogon dystrybuanty: $P(X > x)$ =pbinom(q=x, size=n, prob=p, lower.tail=FALSE);
- funkcja kwantylowa $q(\alpha)$ (kwantyl rzędu $\alpha \in (0,1)$):

  $q(\alpha)$ = qbinom($p=\alpha$, size=n, prob=p, lower.tail=TRUE);

- qbinom($p=\alpha$, size=n, prob=p, lower.tail=FALSE) wyznaczy największą liczbę $c(\alpha)$ spełniającą $P(X > c(\alpha)) \geqslant \alpha$;
- rbinom(n, size, prob) wygeneruje n liczb losowych z rozkładu $binom$(size,prob)

i analogicznie dla pozostałych rozkładów (fragmenty napisane na niebiesko można pominąć).